Metabarcoding and Metagenomics 5: 121–132
DOI 10.3897/mbmg.5.69602

Data Paper

Ciliate SSU-rDNA reference alignments and trees for phylogenetic placements of metabarcoding data

Ľubomír Rajter[1], **Micah Dunthorn**[1,2]

1 *Eukaryotic Microbiology, Faculty of Biology, University of Duisburg-Essen, D-45141, Essen, Germany*
2 *Natural History Museum, University of Oslo, N-0318, Oslo, Norway*

Corresponding author: Ľubomír Rajter (lubomir.rajter@uni-due.de)

Academic editor: Jan Pawlowski | **Received** 1 June 2021 | **Accepted** 11 August 2021 | **Published** 30 August 2021

Abstract

Although ciliates are one of the most dominant microbial eukaryotic groups in many environments, there is a lack of updated global ciliate alignments and reference trees that can be used for phylogenetic placement methods to analyze environmental metabarcoding data. Here we fill this gap by providing reference alignments and trees for those ciliates taxa with available SSU-rDNA sequences derived from identified species. Each alignment contains 478 ciliate and six outgroup taxa, and they were made using different masking strategies for alignment positions (unmasked, masked and masked except the hypervariable V4 region). We constrained the monophyly of the major ciliate groups based on the recently updated classification of protists and based on phylogenomic data. Taxa of uncertain phylogenetic position were kept unconstrained, except for *Mesodinium* species that we constrained to form a clade with the Litostomatea. These ciliate reference alignments and trees can be used to perform taxonomic assignments of metabarcoding data, discover novel ciliate clades, estimate species richness, and overlay measured ecological parameters onto the phylogenetic placements.

Key Words

Ciliophora, diversity, environmental sequencing, HTS, protists, 18S

Introduction

Phylogenetic placement is increasingly used to analyze environmental metabarcoding data. For example, it allows us to retain data that is deeply divergent from taxonomic reference databases (Mahé et al. 2017), perform taxonomic assignments (Bass et al. 2018; Jamy et al. 2020), and quantify the diversity distribution over a reference phylogeny (Barbera et al. 2021). Available phylogenetic placement programs include EPA-NG (Berger et al. 2011; Barbera et al. 2018), PPLACER (Matsen et al. 2010) and APP-SPAM (Blanke and Morgenstern 2020). These programs require a reference alignment and a reference tree. In ciliates, there is a recent reference alignment and tree for just the colpodean ciliates (Rajter et al. 2021) but recent global reference alignments and trees for all ciliates are lacking.

Ciliates are an ancient clade of microbial eukaryotes that originated sometime in the Proterozoic (Dunthorn et al. 2015; Fernandes and Schrago 2019). They are united by their unique cortex arrangement, nuclear dualism, and conjugation as a sexual process (Dunthorn and Katz 2008; Lynn 2008). As ciliates populate most of the aquatic and terrestrial ecosystems on Earth, they are often used to evaluate microbial biodiversity and biogeography (Finlay 2002; Foissner et al. 2009; Dunthorn et al. 2014a; Lentendu et al. 2018; Venter et al. 2018) and for biomonitoring of stressed environments (Stoeck 2018; Dias et al. 2021; Kulaš et al. 2021). The latest classification system of the ciliates recognized eleven main ciliate taxa (Armophorea, Colpodea, Heterotrichea, Karyorelictea, Litostomatea, Nassophorea, Oligohymenophorea, Phyllopharyngea, Plagiopylea, Prostomatea, and Spirotrichea), which are grouped into several higher taxa (Adl et al. 2019) (Fig. 1); additionally there are several *incertae sedis* taxa (*Askenasia*, *Cyclotrichium*, Discotrichidae, Mesodiniidae, *Paraspathidium*, Phacodiniidia, Protocruziidae, and *Pseudotrachelocerca*).

Copyright *L'ubomír Rajtera, Micah Dunthorn.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

There are five global ciliate alignments and trees suitable for phylogenetic placement methods. All are based on the SSU-rDNA, containing the hypervariable V4 and V9 regions that are usually amplified in metabarcoding studies of ciliates. The first was published by Dunthorn et al. (2012b), which contained 175 ciliate sequences representing all 11 major ciliate clades, but outgroups were not included. The second was released a few years later in Dunthorn et al. (2014b), where the number of ciliate sequences was increased to 308, and there were two outgroups. The third and fourth are from Lynn (2008) and Gao et al. (2016), but the alignments and trees were not made publicly available. The fifth is from the EukRef-Ciliophora sequence database by Boscaro et al. (2018), which contained more than 1,200 ciliate sequences (including many unidentified environmental sequences). The identified sequences were not labelled based on the GenBank entries, and sequences with more than 97% similarity were discarded from the dataset. Since the publication of these reference alignments and trees, additional morphologically-identified sequences have been added to GenBank, and the ciliate classification was updated by Adl et al. (2019).

Here we provide updated ciliate reference alignments and trees designed to be used in phylogenetic placement analyses of environmental metabarcoding data. To do this, we: i) created SSU-rDNA dataset containing 478 ingroup and six outgroup sequences, ii) made three multiple sequence alignments with different masking of the ambiguously aligned nucleotide positions, and iii) constructed the maximum likelihood (ML) trees with constrained monophyly of main ciliate taxa based on the Adl et al. (2019) system.

Methods

Dataset with raw sequences

We enriched the ciliate nuclear SSU-rDNA alignment from Dunthorn et al. (2014) with subsequently-published ciliate sequences from GenBank (Clark et al. 2016), to create an updated dataset with 478 ciliate sequences and six outgroup sequences (Suppl. material 20: Table S1). This dataset contained only sequences from morphologically identified species; environmental sequences were not included, except *Cariacothrix* that only has environmental data in GenBank. Redundant sequences representing the same species were excluded, except two *Phacodinium metchnikoffi* populations were retained because this species creates an independent clade with an unknown position in the ciliate tree (Shin et al. 2000). We also included at least one sequence from each high-supported clade (bootstrap value ≥95%) found in trees from the recent ciliate class-specific phylogenetic studies (Suppl. material 1: File S1).

Alignments and masking

Multiple sequence alignments were built with MAFFT v7.471 (Katoh et al. 2019) using an iterative FFT-NS-i algorithm that always re-aligns the alignment after adding a new sequence (Katoh et al. 2002). Sequences were manually trimmed based on the universal eukaryotic SSU-rDNA primers (Medlin et al. 1988) in BioEdit v7.2.5 (Hall 1999). Ambiguous alignment positions were masked using GBLOCKS v0.91b (Talavera and Castresana 2007) with the following block parameters: the minimum number of sequences for a flank position was set at 85%, the maximum number of contiguous non-conserved positions was fixed on eight, minimal block length was set on 10, and we allowed gap positions within the final blocks.

We created three multiple sequence alignments to compare the effects of various masking strategies (Suppl. material 21: Table S2). The first alignment had no masking of ambiguously aligned positions, containing extra phylogenetic information but possibly introducing misleading phylogenetic signal. The second alignment included masking for all ambiguously aligned positions, representing less informative but more biased-free alignment. And the third alignment also had masking for ambiguously aligned positions, except for the hyper-variable V4 region with a relatively high phylogenetic signal (Dunthorn et al. 2012a), representing a trade-off between losing phylogenetic informativeness and potentially biased results; the V4 region is typically used in metabarcoding studies of ciliates. All our reference alignments with corresponding trees can be found in the supplement (Suppl. material 2–Suppl. material 10).

Phylogenetic tree construction and constraints

Phylogenetic trees were inferred from all three alignments using RAXML v8.2.12 (Stamatakis 2014). The GTRGAMMA model was preferably used as it's implemented in RAXML. To find a ML tree among all possible tree structures, we used the SPR tree-rearrangement. This rearrangement algorithm cuts a subpart of the tree and places it in another tree position, creating a new tree structure. We also used the constraint (-g) option in RAXML to force monophyly of all major ciliate subtaxa based on Adl et al. (2019). To compare the constraint's effect on the trees, we made an additional unconstrained tree from the first (unmasked) alignment (Suppl. material 11: File S11, Suppl. material 12: File S12, Suppl. material 13: File S13). For this tree, we performed 1,000 bootstrap replicates to find out the reliability of individual clades. Values <70% were considered low, 70–94% as moderate, and ≥95% as high (Hillis and Bull 1993). Trees were visualized and annotated in ITOL v5.5.1 (Letunic and Bork 2019).

Results and discussion

Phylogenetic description of the reference trees

We forced the monophyly of all ciliate taxa based on the Adl et al. (2019) classification system in our constrained trees. To compare the effect of the constrained option on

the overall topology, we also created an unconstrained tree from the unmasked alignment. The relationships between main ciliate taxa and higher grouping in our trees are described and compared with the literature below; the phylogeny within main taxa is described and discussed in Suppl. material 1: File S1.

All ciliate lineages clustered into two primary groups in all our trees. The first group, containing the Postciliodesmatophora clade, encompassed the Karyorelictea and Heterotrichea. This branching pattern is supported by previous morphological (Lynn 2008) and molecular (Gao and Katz 2014; Gentekaki et al. 2014) studies. Postciliodesmatophora branched off first and had moderate statistical support in our unconstrained tree, while the Intramacronucleata branched off the second but received only low statistical support. The position of Mesodiniidae outside of both main clades in the unconstrained tree is discussed in the *incertae sedis* part below.

The second group, containing the Intramacronucleata clade, encompassed the rest of the major ciliate taxa. Intramacronucleata was further splitting into two clades corresponding to SAL and CONthreeP groups. SAL contained the Spirotrichea, Armophorea (incl. Cariacotrichea, Muranotrichea, and Parablepharismea), and Litostomatea. Although SAL had only low statistical support in our unconstrained tree, its monophyly is highly supported in phylogenomic studies (Gentekaki et al. 2014; Lynn et al. 2018). Relationships inside SAL varied based on the alignment used. Litostomatea was sister to Armophorea+Spirotrichea in both unmasked alignments, while Armophorea was sister to Litostomatea+Spirotrichea in masked and V4 unmasked alignments. But the relationships within SAL are still poorly understood. Morphological and ontogenetic data supporting the Armophorea+Litostomatea grouping were presented in Vďačný et al. (2010). Statistical support from phylogenetic and phylogenomic molecular data is low (Lynn 2008; Gentekaki et al. 2014; Lynn et al. 2018).

The CONthreeP clade is the most diverse, containing six major ciliate groups: Colpodea, Oligohymenoporea, Nassophorea, Phyllopharyngea, Plagiopylea, Prostomatea (Adl et al. 2019). Though the monophyly of CONthreeP had only low statistical support in our unconstrained tree, it is consistently recognized in phylogenetic studies (Lynn 2008) and robustly supported in phylogenomic studies (Lynn et al. 2018). Regarding its intra-relationships, the late branching taxa were consistent in all our trees: Oligohymenophorea was sister to Prostomatea+Plagiopylea. This branching pattern was also recognized in other phylogenetic (e.g., Gao et al. 2016) and phylogenomic (e.g., Maurer-Alcalá et al. 2018) studies. But the relationships varied in the early-branching part of the CONthreeP subtree. Nassophorea+Phyllopharyngea branched off first, and Colpodea branched off the second in the constrained unmasked alignment. Colpodea+Nassophorea branched off first, and Phyllopharyngea branched off the second in masked and V4 unmasked alignments. Finally, Nassophorea branched off first, then Colpodea and Phyllopharyngea in the unconstrained tree.

Constrained monophyly of ciliate lineages

We constrained monophyly of the 11 main ciliate taxa based on the Adl et al. (2019) system that corresponds with a broad consensus from the recent phylogenetic and phylogenomic studies (Lynn 2008; Gentekaki et al. 2017; Lynn et al. 2018) as well as with the morphological view on ciliate phylogeny (Lynn 2003). The classification we used is summarized in Fig. 1, and all taxa with the corresponding classification are listed in Suppl. material 20: Table S1. Below, we discussed some of the taxa that are not strictly monophyletic in all SSU-rDNA phylogenies or based on the phylogenomic data.

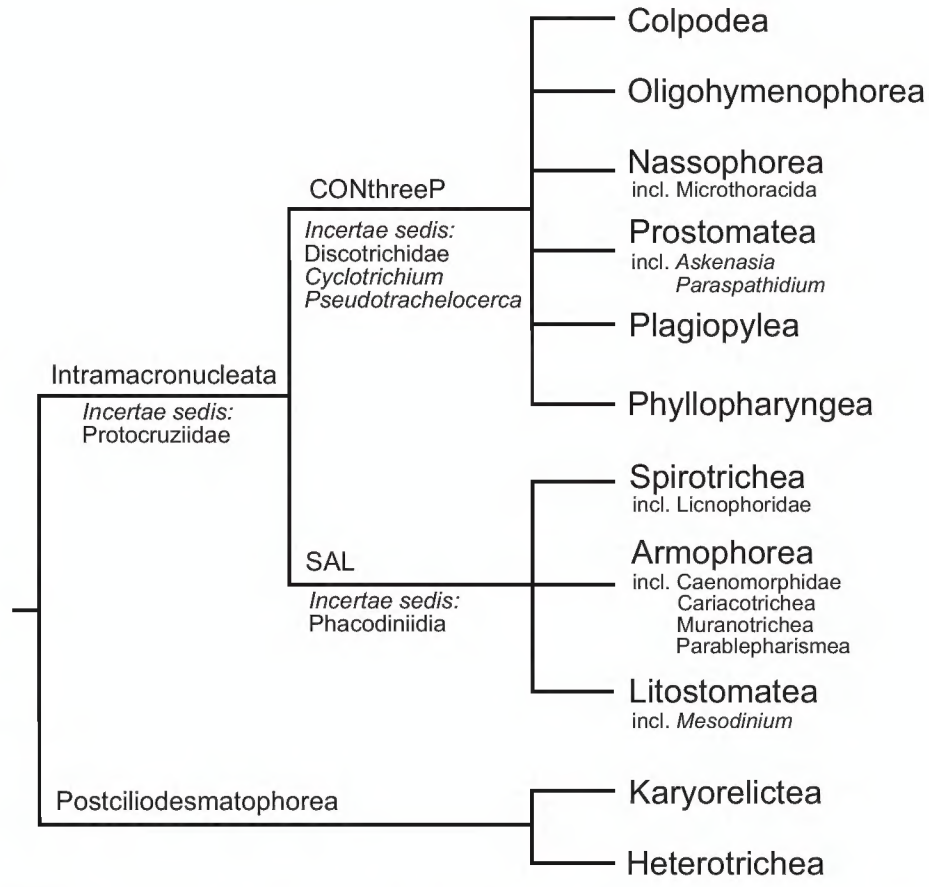

Figure 1. Summary of the new reference trees with the main ciliate taxa and *incertae sedis* groups.

One of these problematic taxa is the Nassophorea. Recent molecular studies have inferred this taxon as non-monophyletic, because its two subtaxa, Microthoracida and Nassulida, have not grouped together (Gong et al. 2009; Fan et al. 2014; Zhang et al. 2014; Pan et al. 2019). But as the definitive phylogenetic position of microthoracids and nassulids has yet to be resolved, we followed the Adl et al. (2019) system here. Thus, we forced Nassophorea to be monophyletic, containing Microthoracida and Nassulida together.

Spirotrichea represents a species-rich ciliate taxon with some taxonomically questionable members (Shin et al. 2000; Lynn and Strüder-Kypke 2002). One of these problematic spirotrichean taxa is the Licnophoridae, containing mostly symbionts of marine animals (Lynn 2008). Licnophoreans were assigned to Spirotrichea because they cluster in its vicinity in SSU-rDNA phylogenies, and because they also possess some typical spirotrichean

morphologies (Lynn and Strüder-Kypke 2002). On the other hand, there are molecular data only for two species, and their phylogenetic position has not been thoroughly evaluated. For example, in our unconstrained tree the family Licnophoridae branched as sister to Armophorea+Spirotrichea, i.e. outside of the class Spirotrichea as in the other studies (Miao et al. 2009; Gao and Katz 2014; Chen et al. 2015). Nevertheless, we follow (Adl et al. 2019) classification here, where licnophoreans are assigned within Spirotrichea, and we forced this grouping in all our constrained trees.

The taxon Armophorea is also non-monophyletic in SSU-rDNA phylogenies (da Silva Paiva et al. 2013; Li et al. 2017; Fernandes et al. 2018; Rotterová et al. 2018), as it contains family Caenomorphidae that is typically branching outside of Armophorea but always somewhere in the Spirotrichea+Armophorea+Litostomatea (SAL) group. In our unconstrained tree, caenomorphids branched as sister to the lineage Litostomatea, as in (Li et al. 2017) and (Rotterová et al. 2018). But as the phylogenetic position of caenomorphids is unstable based on the SSU-rDNA, we kept this family within the Armophorea in all our constrained trees.

The taxon Cariacotrichea was created based on the environmental sequences from the anoxic Cariaco Basin's deep-sea waters in Venezuela (Stoeck et al. 2003b; Edgcomb et al. 2011; Orsi et al. 2012). Cariacotrichea is sometimes referred to as a separate class. Still, as more data are needed for this group, most of the ciliate classification systems considered cariacotricheans as *incertae sedis* clustering somewhere in SAL (Adl et al. 2019). Two other newly established lineages in the vicinity of Armophrea are Parablepharismea and Muranotrichea (Rotterová et al. 2020). In our trees, all three lineages – Cariacotrichea, Parablepharismea, and Muranotrichea – always grouped together creating a sister clade to Armophorea as in Rotterová et al. (2020). For simplicity, we decided to label all these clades as Armophorea in our reference trees, which corresponds to the APM clade defined in Rotterová et al. (2020). For the detailed taxonomical information, please see Suppl. material 20: Table S1.

Although most of our classification followed Adl et al. (2019), an exception is the family Mesodiniidae (*Mesodinium pulex* and *Myrionecta rubra*) that we assigned to the taxon Litostomatea and not as an *incertae sedis* taxon. Mesodiniidae created a separate long-branched clade that clustered between all ciliates and outgroup in our unconstrained tree and other molecular studies (Dunthorn et al. 2014b; Gao et al. 2016). The problematic phylogenetic placement of mesodiniids is caused by the different primary structure of the mesodiniids' SSU-rDNA sequence compared to the typical ciliate SSU-rDNA sequence. But based on the phylogenomic data, Lasek-Nesselquist and Johnson (2019) discovered that mesodiniids have a close affinity to Litostomatea corresponding to the Lynn (2008) classification. They showed that the highly derived mesodiniids' SSU-rDNA sequence results from stolen prey nuclei dominating the cell expression. In light of this new evidence, we decided to constrain the Mesodiniidae to be in the Litostomatea.

Phylogenetic position of some ciliate incertae sedis taxa

Despite the recent advances in sequencing techniques and more molecular data available, the phylogenetic position of some ciliate taxa is still uncertain. We still wait for more data from these enigmatic lineages that we labelled as *incertea sedis* for now. We discuss below those taxa one by one and describe their evolutionary position in our reference trees compared to results from other phylogenetic studies.

Protocruziidae (*Protocruzia adherens* and *Protocruzia contrax*) represented a separate lineage within the Intramacronucleata in all our trees (Fig. 1). Protocruziidae have been formerly assigned to the lineage Spirotrichea (Lynn 2008), but its phylogenetic position is unstable in molecular studies and mainly creates independent clade within Intramacronucleata based on single-gene (Li et al. 2010) and also phylogenomic (Gentekaki et al. 2014) studies. This lineage is placed as *incertae sedis* within Intramacronucleata in the Adl et al. (2019) classification as well.

Phacodiniidia represented by *Phacodinium metchnikoffi* sequences from the Korean and China populations created a separate lineage clustered within SAL lineage as sister to Spirotrichea or as a sister to Litostomatea in our reference trees. Although Phacodiniidia represents a spirotrichean subgroup according to (Lynn 2008), only limited molecular data are available for this taxon, and its systematic positions have never been satisfactorily resolved. Therefore, this taxon is assigned as *incertea sedis* within the SAL group in Adl et al. (2019) and in our reference trees, too (Fig. 1).

Although *Askenasia* sp. had a stable phylogenetic position within the lineage Prostomatea in all our reference trees, this taxon represents *incertea sedis* in Adl et al. (2019). The phylogenetic position of *Askenasia* was thoroughly studied in (Liu et al. 2015), where this genus cluster close to Prostomatea, Plagiopylea, or Oligohymenophorea based on SSU-rDNA phylogeny. But considering morphology, Liu et al. (2015) assumed that *Askenasia* is indeed a prostomatean, as our reference trees also indicated (Fig. 1). The *Paraspathidium apofuscum* sequence was also placed within the lineage Prostomatea in all our reference trees despite this genus having the *incertae sedis* status in Adl et al. (2019).

Cyclotrichium cyclokaryon and *Pseudotrachelocerca trepida* formed a close relationship in all our trees, clustering always inside the CONthreeP group. Both taxa are also *incertae sedis* in Adl et al. (2019) within CONthreeP and grouping near Plagiopylea. Another *incertae sedis* taxon – Discotrichidae – was also placed within the CONthreeP group in all our reference trees. The family Discotrichidae was previously included in the Microthoracida (class Nassophorea) (Lynn 2008), but after inferring its SSU-rDNA phylogeny (Fan et al. 2014; Zhang et al. 2014), this taxon was assigned as *incertae sedis* with-

in CONthreeP (Adl et al. 2019), as our reference trees showed as well (Fig. 1).

How to use our reference alignments and trees in the ciliate diversity research

Phylogenetic placement of metabarcoding data allows researchers to ask broader evolutionary and ecological questions. But without taxonomical expertise on specific taxa, interpretation of these placements is difficult. Here we offer suggestions on some of these interpretations for ciliates using our reference alignments and trees. First, the user can determine the taxonomic position of the investigated environmental sequences as during phylogenetic placement is each sequence attached to the reference tree. The exact phylogenetic positions for all placed sequences can be retrieved using the taxonomical assignment command in the GAPPA package (Czech et al. 2020). In this way, it is possible to taxonomically assign all the investigated environmental sequences, even those that are distantly related to known species and are challenging to assign using standard taxonomic assignment methods (Koski and Golding 2001). The comparison between standard taxonomical assignments and phylogenetic placements is discussed in more detail by Czech and Stamatakis (2019)

Second, the user can discover novel ciliate clades. Despite a long history of the ciliate investigation, novel ciliate clades outside of the main ciliate taxa could be still found in understudied environments such as the anoxic deep-sea waters (Stoeck et al. 2003b; Orsi et al. 2012; Rotterová et al. 2020) or geographical regions such as subtropical and tropical areas (Fernandes et al. 2020). Novel environmental clades within the main ciliate taxa will likely be detected in our reference trees. Many environmental clades have been recovered within almost all main ciliate taxa using a global ciliate tree in Boscaro et al. (2018). Also, numerous unique ciliate environmental sequences have been reported in recent metabarcoding studies from marine and freshwaters habitats (Gimmler et al. 2016; Canals et al. 2020; Fernandes et al. 2020; Zhao 2021). As phylogenetic placements combined with reference alignments will determine the phylogenetic position and the close relatives of the novel environmental clades, molecular ecologists could use this information to assign ecological characteristics of the potential species in the given environmental clade. Based on the environmental sequences, it is possible to create specific phylogenetic probes for fluorescence *in situ* hybridization (FISH) and link morphological and molecular data in a further investigation (Stoeck et al. 2003a; Gimmler and Stoeck 2015).

Third, the user can estimate the total number of ciliate species and their subtaxa. This is possible by using the assign subcommand in the GAPPA package (Czech et al. 2020) that will taxonomically assign placed query sequences into the predefined taxonomy that needs to be provided as an input file. We prepared two different predefine taxonomy input files. The first one contains only assignment to the main ciliate taxa and the outgroup (Suppl. material 14: File S14). Using this input file, it is possible to determine the exact number of placed queries sequences inside and outside of Ciliophora as well as in each main ciliate taxon. The second input file (Suppl. material 15: File S15) includes more detailed system that is based on Adl et al. (2019) classification. Using this input file, it is possible to get detailed taxonomical resolution of the placed query sequences.

Ciliate species diversity has mostly been estimated using morphological differences to define species boundaries. Foissner et al. (2008) estimated that 83–89% of the ciliate diversity is undiscovered and assumed that the total number of ciliate species is between 27,000 and 40,000. On the other hand, Finlay et al. (1998) argued that global ciliate diversity is relatively low, consisting of about 3,000 species. Although recent molecular studies discriminate species based on percentual sequence similarity, these studies often mirror the results obtained from morphological investigations of the same samples (Stoeck et al. 2018; Forster et al. 2019; Pitsch et al. 2019). Molecular ecologists can therefore evaluate the morphological-based estimations of the ciliate species diversity using environmental HTS data combined with our alignments.

Fourth, the user can correlate phylogenetic placements with measured environmental parameters using the edge correlation command in GAPPA. In this way, it is possible to combine environmental parameters with phylogenetic placements and determine which environmental factors influence ciliate community composition.

Conclusions

As the phylogenetic placement methods are increasingly used in metabarcoding studies, here we provided global ciliate reference alignments and trees. We designed and formatted these reference alignments and trees specifically for the ciliate diversity research and phylogenetic placement demands as there are no such available datasets. These files are easily downloadable from the online supplement (Suppl. material 2–Suppl. material 10).

Acknowledgements

We received funding from the Alexander von Humboldt Foundation to LR, and the Deutsche Forschungsgemeinschaft (grant DU1319/5-1) to MD.

References

Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, Campo J del, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov

S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Gall LL, Lynn DH, Mann DG, Massana R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. Journal of Eukaryotic Microbiology 66: 4–119. https://doi.org/10.1111/jeu.12691

Barbera P, Czech L, Lutteropp S, Stamatakis A (2021) SCRAPP: A tool to assess the diversity of microbial samples from phylogenetic placements. Molecular Ecology Resources 21: 340–349. https://doi.org/10.1111/1755-0998.13255

Barbera P, Kozlov AM, Czech L, Morel B, Stamatakis A (2018) EPA-ng: Massively parallel evolutionary placement of genetic sequences. bioRxiv: 291658. https://doi.org/10.1101/291658

Bass D, Czech L, Williams BAP, Berney C, Dunthorn M, Mahé F, Torruella G, Stentiford GD, Williams TA (2018) Clarifying the relationships between Microsporidia and Cryptomycota. Journal of Eukaryotic Microbiology 65: 773–782. https://doi.org/10.1111/jeu.12519

Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Systematic Biology 60: 291–302. https://doi.org/10.1093/sysbio/syr010

Blanke M, Morgenstern B (2020) Phylogenetic placement of short reads without sequence alignment. bioRxiv: 2020.10.19.344986. https://doi.org/10.1101/2020.10.19.344986

Boscaro V, Santoferrara LF, Zhang Q, Gentekaki E, Syberg-Olsen MJ, del Campo J, Keeling PJ (2018) EukRef-Ciliophora: A manually curated, phylogeny-based database of small subunit rRNA gene sequences of ciliates. Environmental Microbiology 20: 2218–2230. https://doi.org/10.1111/1462-2920.14264

Canals O, Obiol A, Muhovic I, Vaqué D, Massana R (2020) Ciliate diversity and distribution across horizontal and vertical scales in the open ocean. Molecular Ecology 29: 2824–2839. https://doi.org/10.1111/mec.15528

Chen X, Ma H, AL-Rasheid K, Miao M (2015) Molecular data suggests the ciliate Mesodinium (Protista: Ciliophora) might represent an undescribed taxon at class level. Zoological Systematics 40: 31–40. https://doi.org/10.11865/zs.20150102

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. Nucleic Acids Research 44: D67–D72. https://doi.org/10.1093/nar/gkv1276

Czech L, Stamatakis A (2019) Scalable methods for analyzing and visualizing phylogenetic placement of metagenomic samples. PLoS ONE 14: e0217050. https://doi.org/10.1371/journal.pone.0217050

Czech L, Barbera P, Stamatakis A (2020) Genesis and Gappa: Processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics 36: 3263–3265. https://doi.org/10.1093/bioinformatics/btaa070

Dias RJP, de Souza PM, Rossi MF, Wieloch AH, da Silva-Neto ID, D'Agosto M (2021) Ciliates as bioindicators of water quality: A case study in the neotropical region and evidence of phylogenetic signals (18S-rDNA). Environmental Pollution 268: 115760. https://doi.org/10.1016/j.envpol.2020.115760

Dunthorn M, Katz L (2008) Richness of morphological hypotheses in ciliate systematics allows for detailed assessment of homology and comparisons with gene trees. Denisia 23: 389–394. http://archive.org/details/denisia-23-389-394 [April 10, 2021]

Dunthorn M, Klier J, Bunge J, Stoeck T (2012a) Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. Journal of Eukaryotic Microbiology 59: 185–187. https://doi.org/10.1111/j.1550-7408.2011.00602.x

Dunthorn M, Stoeck T, Wolf K, Breiner H-W, Foissner W (2012b) Diversity and endemism of ciliates inhabiting Neotropical phytotelmata. Systematics and Biodiversity 10: 195–205. https://doi.org/10.1080/14772000.2012.685195

Dunthorn M, Stoeck T, Clamp J, Warren A, Mahé F (2014a) Ciliates and the rare biosphere: A review. Journal of Eukaryotic Microbiology 61: 404–409. https://doi.org/10.1111/jeu.12121

Dunthorn M, Otto J, Berger SA, Stamatakis A, Mahé F, Romac S, de Vargas C, Audic S, Consortium B, Stock A, Kauff F, Stoeck T (2014b) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. Molecular Biology and Evolution 31: 993–1009. https://doi.org/10.1093/molbev/msu055

Dunthorn M, Lipps JH, Dolan JR, Saab MA-A, Aescht E, Bachy C, de Cao MSB, Berger H, Bourland WA, Choi JK, Clamp J, Doherty M, Gao F, Gentekaki E, Gong J, Hu X, Huang J, Kamiyama T, Johnson MD, Kammerlander B, Kim SY, Kim Y-O, la Terza A, Laval-Peuto M, Lipscomb D, Lobban CS, Long H, Luporini P, Lynn DH, Macek M, Mansergh RI, Martín-Cereceda M, McManus GG, Montagnes DJS, Ong'ondo GO, Patterson DJ, Pérez-Uz B, Quintela-Alonso P, Safi LSL, Santoferrara LF, Sonntag B, Song W, Stoeck T, Stoecker DK, Strüder-Kypke MC, Trautmann I, Utz LRP, Vallesi A, Vd'ačný P, Warren A, Weisse T, Wickham SA, Yi Z, Zhang W, Zhan Z, Zufall R, Agatha S (2015) Ciliates – Protists with complex morphologies and ambiguous early fossil record. Marine Micropaleontology 119: 1–6. https://doi.org/10.1016/j.marmicro.2015.05.004

Edgcomb V, Orsi W, Taylor GT, Vdacny P, Taylor C, Suarez P, Epstein S (2011) Accessing marine protists from the anoxic Cariaco Basin. The ISME Journal 5: 1237–1241. https://doi.org/10.1038/ismej.2011.10

Fan X, Pan H, Li L, Jiang J, Al-Rasheid KAS, Gu F (2014) Phylogeny of the poorly known ciliates, Microthoracida, a systematically confused taxon (Ciliophora), with morphological reports of three species. Journal of Eukaryotic Microbiology 61: 227–237. https://doi.org/10.1111/jeu.12099

Fernandes NM, Schrago CG (2019) A multigene timescale and diversification dynamics of Ciliophora evolution. Molecular Phylogenetics and Evolution 139: 106521. https://doi.org/10.1016/j.ympev.2019.106521

Fernandes NM, Campello-Nunes PH, Paiva TS, Soraes CAG, Silva-Neto ID (2020) Ciliate diversity from aquatic environments in the Brazilian Atlantic Forest as revealed by high-throughput DNA sequencing. Microbial Ecology 81: 630–643. https://doi.org/10.1007/s00248-020-01612-8

Fernandes NM, Vizzoni VF, Borges B do N, Soares CAG, da Silva-Neto ID, Paiva T da S (2018) Molecular phylogeny and comparative morphology indicate that odontostomatids (Alveolata, Ciliophora) form a distinct class-level taxon related to Armophorea. Molecular Phylogenetics and Evolution 126: 382–389. https://doi.org/10.1016/j.ympev.2018.04.026

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. Science (New York, N.Y.) 296: 1061–1063. https://doi.org/10.1126/science.1070710

Finlay BJ, Esteban GF, Fenchel T (1998) Protozoan diversity: Converging estimates of the global number of free-living ciliate species. Protist 149: 29–37. https://doi.org/10.1016/S1434-4610(98)70007-0

Foissner W, Chao A, Katz LA (2008) Diversity and geographic distribution of ciliates (Protista: Ciliophora). Biodiversity and Conservation 17: 345–363. https://doi.org/10.1007/s10531-007-9254-7

Foissner W, Chao A, Katz LA (2009) Diversity and geographic distribution of ciliates (Protista: Ciliophora). In: Foissner W, Hawksworth DL (Eds) Protist Diversity and Geographical Distribution. Topics in Biodiversity and Conservation. Springer Netherlands, Dordrecht, 111–129. https://doi.org/10.1007/978-90-481-2801-3_9

Forster D, Filker S, Kochems R, Breiner H-W, Cordier T, Pawlowski J, Stoeck T (2019) A comparison of different ciliate metabarcode genes as bioindicators for environmental impact assessments of salmon aquaculture. Journal of Eukaryotic Microbiology 66: 294–308. https://doi.org/10.1111/jeu.12670

Gao F, Katz LA (2014) Phylogenomic analyses support the bifurcation of ciliates into two major clades that differ in properties of nuclear division. Molecular Phylogenetics and Evolution 70: 240–243. https://doi.org/10.1016/j.ympev.2013.10.001

Gao F, Warren A, Zhang Q, Gong J, Miao M, Sun P, Xu D, Huang J, Yi Z, Song W (2016) The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Scientific Reports 6: e24874. https://doi.org/10.1038/srep24874

Gentekaki E, Kolisko M, Gong Y, Lynn D (2017) Phylogenomics solves a long-standing evolutionary puzzle in the ciliate world: The subclass Peritrichia is monophyletic. Molecular Phylogenetics and Evolution 106: 1–5. https://doi.org/10.1016/j.ympev.2016.09.016

Gentekaki E, Kolisko M, Boscaro V, Bright KJ, Dini F, Di Giuseppe G, Gong Y, Miceli C, Modeo L, Molestina RE, Petroni G, Pucciarelli S, Roger AJ, Strom SL, Lynn DH (2014) Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon Protocruzia and unravels the deep phylogenetic affinities of the ciliate lineages. Molecular Phylogenetics and Evolution 78: 36–42. https://doi.org/10.1016/j.ympev.2014.04.020

Gimmler A, Stoeck T (2015) Mining environmental high-throughput sequence data sets to identify divergent amplicon clusters for phylogenetic reconstruction and morphotype visualization: EnvHTS data mining. Environmental Microbiology Reports 7: 679–686. https://doi.org/10.1111/1758-2229.12307

Gimmler A, Korn R, de Vargas C, Audic S, Stoeck T (2016) The Tara Oceans voyage reveals global diversity and distribution patterns of marine planktonic ciliates. Scientific Reports 6: e33555. https://doi.org/10.1038/srep33555

Gong J, Stoeck T, Yi Z, Miao M, Zhang Q, Roberts DM, Warren A, Song W (2009) Small subunit rRNA phylogenies show that the class nassophorea is not monophyletic (Phylum Ciliophora). The Journal of Eukaryotic Microbiology 56: 339–347. https://doi.org/10.1111/j.1550-7408.2009.00413.x

Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic snalysis. Systematic Biology 42: 182–192. https://doi.org/10.1093/sysbio/42.2.182

Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F (2020) Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. Molecular Ecology Resources 20: 429–443. https://doi.org/10.1111/1755-0998.13117

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20: 1160–1166. https://doi.org/10.1093/bib/bbx108

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30: 3059–3066. https://doi.org/10.1093/nar/gkf436

Koski LB, Golding GB (2001) The closest BLAST hit is often not the nearest neighbor. Journal of Molecular Evolution 52: 540–542. https://doi.org/10.1007/s002390010184

Kulaš A, Gulin V, Matoničkin Kepčija R, Žutinić P, Sertić Perić M, Orlić S, Kajan K, Stoeck T, Lentendu G, Čanjevac I, Martinić I, Gligora Udovič M (2021) Ciliates (Alveolata, Ciliophora) as bioindicators of environmental pressure: A karstic river case. Ecological Indicators 124: 107430. https://doi.org/10.1016/j.ecolind.2021.107430

Lasek-Nesselquist E, Johnson MD (2019) A phylogenomic approach to clarifying the relationship of Mesodinium within the Ciliophora: A case study in the complexity of mixed-species transcriptome analyses. Zufall R (Ed.). Genome Biology and Evolution 11: 3218–3232. https://doi.org/10.1093/gbe/evz233

Lentendu G, Mahé F, Bass D, Rueckert S, Stoeck T, Dunthorn M (2018) Consistent patterns of high alpha and low beta diversity in tropical parasitic and free-living protists. Molecular Ecology 27: 2846–2857. https://doi.org/10.1111/mec.14731

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research 47: W256–W259. https://doi.org/10.1093/nar/gkz239

Li L, Thorsten S, Kyoon SM, Al-Rasheid Khaled AS, Al-Khedhairy Bdulaziz A, Song W (2010) Protocruzia, a highly ambiguous ciliate (Protozoa; Ciliophora): Very likely an ancestral form for Heterotrichea, Colpodea or Spirotrichea? With reevaluation of its evolutionary position based on multigene analyses. Science China. Life Sciences 53: 131–138. https://doi.org/10.1007/s11427-010-0012-9

Li S, Bourland WA, Al-Farraj SA, Li L, Hu X (2017) Description of two species of caenomorphid ciliates (Ciliophora, Armophorea): Morphology and molecular phylogeny. European Journal of Protistology 61: 29–40. https://doi.org/10.1016/j.ejop.2017.08.001

Liu A, Yi Z, Lin X, Hu X, Al-Farraj SA, Al-Rasheid KAS (2015) Molecular phylogenetic lineage of Plagiopogon and Askenasia (Protozoa, Ciliophora) revealed by their gene sequences. Journal of Ocean University of China 14: 724–730. https://doi.org/10.1007/s11802-015-2559-3

Lynn D (2008) The ciliated protozoa: Characterization, classification, and guide to the literature. 3rd edn. Springer Netherlands, 605 pp. https://doi.org/10.1007/978-1-4020-8239-9

Lynn DH (2003) Morphology or molecules: How do we identify the major lineages of ciliates (Phylum Ciliophora). European Journal of Protistology 39: 356–364. https://doi.org/10.1078/0932-4739-00004

Lynn DH, Strüder-Kypke M (2002) Phylogenetic position of Licnophora, Lechriopyla, and Schizocaryum, three unusual ciliates (phylum Ciliophora) endosymbiotic in echinoderms (phylum Echinodermata). The Journal of Eukaryotic Microbiology 49: 460–468. https://doi.org/10.1111/j.1550-7408.2002.tb00229.x

Lynn DH, Kolisko M, Bourland W (2018) Phylogenomic analysis of Nassula variabilis n. sp., Furgasonia blochmanni, and Pseudomicrothorax dubius confirms a nassophorean clade. Protist 169: 180–189. https://doi.org/10.1016/j.protis.2018.02.002

Mahé F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell EAD, Seppey CVW, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nature Ecology & Evolution 1: 1–8. https://doi.org/10.1038/s41559-017-0091

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11: e538. https://doi.org/10.1186/1471-2105-11-538

Maurer-Alcalá XX, Yan Y, Pilling OA, Knight R, Katz LA (2018) Twisted tales: Insights into genome diversity of ciliates using single-cell 'omics. Genome Biology and Evolution 10: 1927–1939. https://doi.org/10.1093/gbe/evy133

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71: 491–499. https://doi.org/10.1016/0378-1119(88)90066-2

Miao M, Shao C, Jiang J, Li L, Stoeck T, Song W (2009) Caryotricha minuta (Xu et al., 2008) nov. comb., a unique marine ciliate (Protista, Ciliophora, Spirotrichea), with phylogenetic analysis of the ambiguous genus Caryotricha inferred from the small-subunit rRNA gene sequence. International Journal of Systematic and Evolutionary Microbiology 59: 430–438. https://doi.org/10.1099/ijs.0.65855-0

Orsi W, Edgcomb V, Faria J, Foissner W, Fowle WH, Hohmann T, Suarez P, Taylor C, Taylor GT, Vďačný P, Epstein SS (2012) Class Cariacotrichea, a novel ciliate taxon from the anoxic Cariaco Basin, Venezuela. International Journal of Systematic and Evolutionary Microbiology 62: 1425–1433. https://doi.org/10.1099/ijs.0.034710-0

Pan B, Chen X, Hou L, Zhang Q, Qu Z, Warren A, Miao M (2019) Comparative genomics analysis of ciliates provides insights on the evolutionary history within "Nassophorea–Synhymenia–Phyllopharyngea" assemblage. Frontiers in Microbiology 10: e2819. https://doi.org/10.3389/fmicb.2019.02819

Pitsch G, Bruni EP, Forster D, Qu Z, Sonntag B, Stoeck T, Posch T (2019) Seasonality of planktonic freshwater ciliates: Are analyses based on V9 regions of the 18S rRNA gene correlated with morphospecies counts? Frontiers in Microbiology 10: e248. https://doi.org/10.3389/fmicb.2019.00248

Rajter Ľ, Ewers I, Graupner N, Vďačný P, Dunthorn M (2021) Colpodean ciliate phylogeny and reference alignments for phylogenetic placements. European Journal of Protistology 77: 125747. https://doi.org/10.1016/j.ejop.2020.125747

Rotterová J, Bourland W, Čepička I (2018) Tropidoatractidae fam. nov., a deep branching lineage of Metopida (Armophorea, Ciliophora) found in diverse habitats and possessing prokaryotic symbionts. Protist 169: 362–405. https://doi.org/10.1016/j.protis.2018.04.003

Rotterová J, Salomaki E, Pánek T, Bourland W, Žihala D, Táborský P, Edgcomb VP, Beinart RA, Kolísko M, Čepička I (2020) Genomics of new ciliate lineages provides insight into the evolution of obligate anaerobiosis. Current Biology 30: 2037–2050.e6. https://doi.org/10.1016/j.cub.2020.03.064

Shin MK, Hwang UW, Kim W, Wright A-DG, Krawczyk C, Lynn DH (2000) Phylogenetic position of the ciliates Phacodinium (Order Phacodiniida) and Protocruzia (Subclass Protocruziidia) and systematics of the spirotrich ciliates examined by small subunit ribosomal RNA gene sequences. European Journal of Protistology 36: 293–302. https://doi.org/10.1016/S0932-4739(00)80005-X

da Silva Paiva T, do Nascimento Borges B, da Silva-Neto ID (2013) Phylogenetic study of class Armophorea (Alveolata, Ciliophora) based on 18S-rDNA data. Genetics and Molecular Biology 36: 571–585. https://doi.org/10.1590/S1415-47572013000400017

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stoeck T (2018) Metabarcoding of benthic ciliate communities shows high potential for environmental monitoring in salmon aquaculture. Ecological Indicators 12.

Stoeck T, Fowle WH, Epstein SS (2003a) Methodology of protistan discovery: From rRNA detection to quality scanning electron microscope images. Applied and Environmental Microbiology 69: 6856–6863. https://doi.org/10.1128/AEM.69.11.6856-6863.2003

Stoeck T, Taylor GT, Epstein SS (2003b) Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). Applied and Environmental Microbiology 69: 5656–5663. https://doi.org/10.1128/AEM.69.9.5656-5663.2003

Stoeck T, Kochems R, Forster D, Lejzerowicz F, Pawlowski J (2018) Metabarcoding of benthic ciliate communities shows high potential for environmental monitoring in salmon aquaculture. Ecological Indicators 85: 153–164. https://doi.org/10.1016/j.ecolind.2017.10.041

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56: 564–577. https://doi.org/10.1080/10635150701472164

Vďačný P, Orsi W, Foissner W (2010) Molecular and morphological evidence for a sister group relationship of the classes Armophorea and Litostomatea (Ciliophora, Intramacronucleata, Lamellicorticata infraphyl. nov.), with an account on basal litostomateans. European Journal of Protistology 46: 298–309. https://doi.org/10.1016/j.ejop.2010.07.002

Venter PC, Nitsche F, Scherwass A, Arndt H (2018) Discrepancies between molecular and morphological databases of soil ciliates studied for temperate grasslands of Central Europe. Protist 169: 521–538 https://doi.org/10.1016/j.protis.2018.04.001.

Zhang Q, Yi Z, Fan X, Warren A, Gong J, Song W (2014) Further insights into the phylogeny of two ciliate classes Nassophorea and Prostomatea (Protista, Ciliophora). Molecular Phylogenetics and Evolution 70: 162–170. https://doi.org/10.1016/j.ympev.2013.09.015

Zhao F (2021) Bathymetric gradient shapes the community composition rather than the species richness of deep-sea benthic ciliates. Science of the Total Environment 11. https://doi.org/10.1016/j.scitotenv.2020.142623

Supplementary material 1
File S1

Author: Ľubomír Rajter
Data type: pdf file
Explanation note: Description of the relationships within the main ciliate lineages in the reference trees.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl1

Supplementary material 2
File S2
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Masked reference alignment except the V4 region in the FASTA format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl2

Supplementary material 3
File S3
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Masked reference alignment except the V4 region in the PHYLIP format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl3

Supplementary material 4
File S4
Author: Ľubomír Rajter
Data type: NEWICK tree
Explanation note: Reference tree built from the masked alignment except the V4 region in the NEWICK format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl4

Supplementary material 5
File S5
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Unmasked reference alignment in the FASTA format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl5

Supplementary material 6
File S6
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Unmasked reference alignment in the PHYLIP format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl6

Supplementary material 7
File S7
Author: Ľubomír Rajter
Data type: NEWICK tree
Explanation note: Reference tree built from the unmasked alignment in the NEWICK format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl7

Supplementary material 8
File S8
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Masked reference alignment in the FASTA format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl8

Supplementary material 9
File S9
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Masked reference alignment in the PHYLIP format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl9

Supplementary material 10
File S10
Author: Ľubomír Rajter
Data type: NEWICK tree
Explanation note: Reference tree built from the masked alignment in the NEWICK format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl10

Supplementary material 11
File S11
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Unconstrained unmasked alignment in the FASTA format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl11

Supplementary material 12
File S12
Author: Ľubomír Rajter
Data type: multiple sequence alignment
Explanation note: Unconstrained unmasked alignment in the PHYLIP format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl12

Supplementary material 13
File S13
Author: Ľubomír Rajter
Data type: NEWICK tree
Explanation note: Phylogenetic tree built from the unconstrained unmasked alignment in the NEWICK format.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl13

Supplementary material 14
File S14
Author: Ľubomír Rajter
Data type: text file
Explanation note: Input file for a taxonomical assignment function in the GAPPA package; this file containing a tab-separated list of reference taxa with corresponding taxonomical assignment to the main ciliate lineages or *incertae sedis*.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl14

Supplementary material 15
File S15
Author: Ľubomír Rajter
Data type: text file
Explanation note: Input file for a taxonomical assignment function in the GAPPA package; this file containing a tab-separated list of reference taxa with corresponding taxonomical assignment.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl15

Supplementary material 16
Figure S1
Author: Ľubomír Rajter
Data type: figure (pdf file)
Explanation note: Reference tree built from the masked alignment except the V4 region.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl16

Supplementary material 17
Figure S2
Author: Ľubomír Rajter
Data type: figure (pdf file)
Explanation note: Reference tree built from the unmasked alignment.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl17

Supplementary material 18
Figure S3
Author: Ľubomír Rajter
Data type: figure (pdf file)
Explanation note: Reference tree built from the masked alignment.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl18

Supplementary material 19
Figure S4
Author: Ľubomír Rajter
Data type: figure (pdf file)
Explanation note: Reference tree built from the unconstrained unmasked alignment.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl19

Supplementary material 20
Table S1
Author: Ľubomír Rajter
Data type: word document
Explanation note: Complete list of the taxa with the sequence entries used in this study.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl20

Supplementary material 21
Table S2
Author: Ľubomír Rajter
Data type: word document
Explanation note: Characterization and evolutionary models of the alignments analyzed.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.69602.suppl21